

REMARKS

This communication is in response to the Office Action mailed on September 13, 2007. In the Office Action, claims 1-31 were pending of which claims 1-31 were rejected.

Information Disclosure Statement

The Office Action reports that the Information Disclosure Statement (IDS) filed on 8/17/07 failed to comply with 37 C.F.R. 1.98(a)(3). The previous Office Action included the same objection to the IDS filed on 8/17/07. Applicants explained in the previous amendment that they believed that a statement of relevance from the individual designated in 37 C.F.R. 1.56(c) had been provided as follows:

This paper is an overview of the state-of-the-art of methods for the Chinese word segmentation task, in particular some investigations of overlapping ambiguity distribution in the corpus, and the overlapping ambiguity detection coverage of the FMM+BMM method.

It is respectfully submitted that this statement is the necessary explanation of relevance in accordance with 37 C.F.R. 1.98(a)(3)(i). In the current Office Action the examiner further requests an English translation of the reference. It is understood that an English translation in accordance with 37 C.F.R. 1.98(a)(3)(ii) is to be provided if it is "within the possession, custody, or control of, or is readily available to any individual designated in 37 C.F.R. 1.56(c)." In the previous amendment, it was explained that the inventors are Chinese and that the reference was published in Chinese. It is believed an English translation is not within the possession, custody or control of any individual designated 37 C.F.R. 1.56(c) in the instant case. It is also believed that applicants do not have an affirmative duty to translate foreign language references into English to comply with the requirements of 37 C.F.R. 1.98. Therefore, it is respectfully requested that the above objection to the IDS be

withdrawn.

Rejections based on 35 U.S.C. §103

The Office Action next reports that claims 1-4, 6-7, 14-21, 23, 25-26, and 28 were rejected under 35 U.S.C. §103 as being unpatentable over U.S. Patent No. 5,806,021 to Chen et al. (hereinafter "Chen") in view of U.S. Patent No. 6,968,308 to Brockett et al. (hereinafter "Brockett") It is respectfully submitted that the cited references even when combined do not teach or suggest all of the features of claim 1.

The examiner states in "Response to Arguments" that applicants had argued that Chen in view of Brockett do not teach the feature of processing a sentence of Chinese characters into constituent words. Applicants respectfully assert that applicants' arguments were more comprehensive than is herein described; and therefore, do not acquiesce to the description. Processing a sentence of Chinese characters into constituent words is believed to be a preliminary step common to many methods of Chinese language text processing. However, the present inventions as recited in the present claims relate to resolving overlapping ambiguity strings of Chinese characters where processing Chinese characters into constituent words is only a preliminary but not critical step.

Without admitting that the cited combination reads on the previous presented claim, claim 1 has been amended for further clarification. Claim 1 recites a computer readable storage media storing instructions readable by a computer which, when implemented, cause the computer to perform a method comprising: segmenting a sentence of Chinese characters into constituent Chinese words having one or more Chinese characters; recognizing an overlapping ambiguity string in the segmented sentence, wherein the overlapping ambiguity string comprises at least three Chinese characters having at least two possible segmentations, wherein each possible segmentation comprises a right portion and a left

portion; obtaining probability information for each possible segmentation, wherein the probability information is based on at least one context feature adjacent the overlapping ambiguity string and one of the left portion or the right portion of the possible segmentation, wherein the at least one context feature comprises a Chinese character; and outputting an indication for selecting one of the at least two possible segmentations as a function of the obtained probability information. [emphasis added]

The amendments to claim 1 further clarify that an overlapping ambiguity string has at least two possible segmentations and each of the possible segmentations comprises a left portion and a right portion. The left and right portions each necessarily has at least one Chinese character since the overlapping ambiguity string has at least three characters. Thus if an overlapping ambiguity string has three characters "ABC" where A, B, and C are Chinese characters then the two possible segmentations are AB/C and AB/C. In this case, one possible segmentation, AB/C, would have AB as a left portion and C as a right portion. The other possible segmentation would have A as a left portion and BC as a right portion. Claim 1 further clarifies that the probability information of a possible segmentation is based on at least one adjacent context feature and a left or right portion of the possible segmentation.

It is submitted that the cited combination Chen in view of Brockett do not teach or suggest all of the features of claim 1. The Office Action cites Chen as the primary reference. Chen discloses a word segmenter or breaker that performs continuous segmentation of text using at least two approaches. One approach employs Forward-Backward Maximum Matching where the segmentation (either forward or backward) is selected based on the likelihood. Another approach is a statistical stack method, which is slower but more accurate than the first approach. This second approach can be selected if accuracy is the primary concern and not speed.

It is believed that Chen is directed towards continuous segmentation of Chinese text but nowhere discloses a method of resolving overlapping ambiguity strings in Chinese.

The Office Action does admit that Chen does not disclose recognizing an overlapping ambiguity string in the input sentence, wherein the overlapping ambiguity string comprises at least three Chinese characters having at least two possible segmentations. However, overlapping ambiguity strings are a major cause of segmentation errors of Chinese text and accurate selection of the correct segmentation of overlapping ambiguity strings was described as a major purpose of the present inventions as recited in the pending claims. Thus, the features of claim 1 relating to overlapping ambiguity strings (that are absent in Chen) are critical to the inventions as recited in claim 1. Therefore, it is not well understood why Chen is selected as the primary reference. Thus, applicants respectfully request that Chen be withdrawn as the primary reference.

Brockett discloses a method of segmenting non-segmented text using syntactic parse. However, it is believed that Brockett is related to the Japanese language, which uses four different kinds of script including kanji, haragana, katakana, and roma. These four scripts can be used to spell the same word, which results in orthographic variations of the word. Chinese language is generally known as "kanji" in the Japanese language and is only one script used in Japanese. It is believed that Brockett discloses a method of segmentation that accounts for these orthographic variations. It is true that Brockett does mention that Chinese is an unsegmented language as is Japanese and Korean. However, it is believed that Brockett cannot be modified for use with Chinese because Chinese does not have the orthographic variations that Brockett is designed to account for in its segmentation method.

The Office Action does refer to Col. 6, lines 6-42 of

Brockett as disclosing obtaining probability information as recited in claim 1. However, on further inspection it is believed that Brockett is describing something very different than the features of claim 1. Brockett describes a word breaker that searches for words in a data structure known as a "trie" where words are not listed sequentially but are instead represented by chains of states. It is believed that this section of Brockett has little or nothing to do with obtaining probability information as recited in claim 1.

In view of the foregoing, it is believed that claim 1 is patentable over the cited art. Claims 2-12 depend on claim 1 and are believed to be separately patentable. Reconsideration and allowance of claims 1-12 are respectfully requested.

Independent claim 14 was also rejected based on the same combination of Chen and Brockett. Without admitting that the cited combination reads on the previously presented claim, claim 14 has been amended to recite a method of segmentation of a sentence of Chinese text, the sentence having an overlapping ambiguity string, the method comprising: generating a Forward Maximum Matching (FMM) segmentation of the sentence; generating a Backward Maximum Matching (BMM) segmentation of the sentence; recognizing the overlapping ambiguity string based on a difference between the FMM segmentation and the BMM segmentation; obtaining probability information based on at least one context feature surrounding the overlapping ambiguity string and at least part of the overlapping ambiguity string, wherein the at least one context feature comprises a Chinese character; and outputting an indication for selecting one of the FMM segmentation and the BMM segmentation as a function of obtained probability information.  
[emphasis added]

Claim 14 has also been amended in a manner similar and is similar in scope to claim 1. Thus, the remarks above are hereby incorporated by reference. Claim 14 now clarifies the at least one

context feature is a Chinese character. The obtained probability information is based on at least one context feature and part of the segmentation of the overlapping ambiguity string. The probability information is then used to select one of the FMM or BMM segmentation of the overlapping ambiguity string.

In light of the foregoing, it is believed that the cited combination does not teach or suggest all of the features of claim 14. Thus, claim 14 is believed to be patentable over the cited art. Claims 15-24 depend on claim 14 and are believed to be separately patentable. Reconsideration and allowance of claims 14-24 are respectfully requested.

The Office Action further cites the same combination against independent claim 25. Claim 25 has been amended to recite a method of segmenting a sentence of Chinese text comprising: recognizing an overlapping ambiguity string in the sentence; receiving probability information from an N-gram language model comprising probability information for constituent words of the overlapping ambiguity string and context features surrounding the overlapping ambiguity string, wherein the context features comprise at least one Chinese character; resolving the overlapping ambiguity string based on the received probability information. [emphasis added]

The discussion of the cited references is hereby incorporated by reference. Claim 25 has been amended so that the received probability information from the N-gram language model is based on constituent words and at least one context feature or Chinese character surrounding the overlapping ambiguity string. As discussed above, it is submitted that the cited references do not teach or suggest all of the features of claim 25.

In light of the foregoing, it is believed that claim 25 is patentable over the cited art. Claims 26-31 depend on claim 25 and are believed to be separately patentable. Reconsideration and allowance of claims 25-31 is respectfully requested.

The foregoing remarks are intended to assist the Office in examining the application and in the course of explanation may employ shortened or more specific or variant descriptions of some of the claim language. Such descriptions are not intended to limit the scope of the claims; the actual claim language should be considered in each case. Furthermore, the remarks are not to be considered exhaustive of the facets of the invention which are rendered patentable, being only examples of certain advantageous features and differences, which applicant's attorney chooses to mention at this time. For the foregoing reasons, applicant reserves the right to submit additional evidence showing the distinction between applicant's invention to be unobvious in view of the prior art.

Furthermore, in commenting on the references and in order to facilitate a better understanding of the differences that are expressed in the claims, certain details of distinction between the same and the present invention have been mentioned, even though such differences do not appear in all of the claims. It is not intended by mentioning any such unclaimed distinctions to create any implied limitations in the claims.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: /Linda Ji, Reg. No. 49,027/

Linda Ji, Reg. No. 49,027  
900 Second Avenue South, Suite 1400  
Minneapolis, Minnesota 55402-3319  
Phone: (612) 334-3222 Fax: (612) 334-3312